Research Policy 53 (2024) 104971

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Linguistic distance to English impedes research performance

Yihui Cao [a,*], Robin C. Sickles [b], Thomas P. Triebs [c], Justin Tumlinson [d,e,1]

[a] University of Westminster, 35 Marylebone Rd, London NW1 5LS, United Kingdom
[b] Rice University, 6100 Main St, Houston, TX 77005, United States of America
[c] Loughborough University, Epinal Way, Loughborough LE11 3TU, United Kingdom
[d] University of Exeter, Stocker Rd, Exeter EX4 4PY, United Kingdom
[e] Technical University of Munich, Arcisstraße 21, Munich 80333, Germany

ARTICLE INFO

Keywords:
Language
Linguistic distance
English
University research performance

ABSTRACT

Today, scientific knowledge is predominantly disseminated in English. We show that global universities' research performance, as measured by publications in top journals, declines as the differences between their local language and English increase. This effect is robust to controls for university factors like proportion of international staff and faculty-to-student ratio, as well as country-level factors like economic development, youth academic achievement, university degree rate, politics, culture, trade with and geographic distance to English-speaking countries, among others. This quantification of the research performance penalties induced by linguistic distance from the *lingua franca* may inform policy makers who must balance trade-offs between embracing English against cultural and local labor market pressures to orient around the local language.

1. Introduction

Universities' research performance matters. They generate economic benefits and alleviate social problems through commercialized inventions and university technology transfers (Hermanu et al., 2022; Rosenberg and Nelson, 1994; Conti and Gaule, 2011). Universities benefit countries' innovation systems and firms' innovation strategies (Giannopoulou et al., 2019; Bercovitz and Feldman, 2007). They contribute to knowledge accumulation and regional economic development (Metcalfe, 2010). Universities play an important role in broadening international knowledge networks and talent mobility (Baruffaldi and Landoni, 2016). University research significantly increases corporate patents, promotes research and development projects, and indirectly stimulates local innovation (Jaffe, 1989; Geuna and Nesta, 2006; Tether and Tajar, 2008; Cohen et al., 2002; Kim et al., 2005). Startups apply universities' research to commercial innovation (Laursen and Salter, 2004).

English increasingly dominates higher education and international scientific communication. From 1880 to 1980, the proportion of scientific publications in English rose from 36% to 64%, and in the last decade to more than 90% (Hamel, 2007; Ammon, 2012). Many international journals have adopted an English-only policy, but even those without such policy often encourage scholars to submit manuscripts in English (Di Bitetti and Ferreras, 2017). While research potential may be distributed uniformly around the globe, English proficiency is not.

We show that having a native language similar to English conveys an economically important and statistically significant research advantage. In particular, the greater the linguistic difference between non-English speaking regions' local language and English (henceforth, linguistic distance or LD), the worse their universities' research performance, as captured by a subject-level research performance indicator from Shanghai Ranking's *Global Ranking of Academic Subjects* (GRAS).

Fig. 1 scatter-plots the linguistic distance of the native language of our sample universities' location to English (horizontal axis) against their 2020 GRAS average Q1 score across social sciences subjects.[2] Q1 score indicates the number of papers published by a university in journals with first quartile impact factors within an academic subject (Shanghai Ranking, 2020). Linguistic distance is a normalized measure from 0 to 100, larger scores indicating greater genetic difference from English (eLinguistics, 2020). Within a single set of national institutions, Belgium's universities (highlighted) in the Dutch ($LD = 27.20$, average social science $Q1 = 42.1$) speaking region have 57% higher social

* Corresponding author.
E-mail addresses: y.cao@westminster.ac.uk (Y. Cao), rsickles@rice.edu (R.C. Sickles), t.triebs@lboro.ac.uk (T.P. Triebs), j.tumlinson@exeter.ac.uk (J. Tumlinson).
1 Authors are listed alphabetically.

2 For visual clarity, we limit this scatterplot to social sciences, where the effect is most easily seen in an uncontrolled linear regression, and we aggregate these Q1 scores to the university average across social sciences subjects. Our controlled log–log regressions are across all subjects.

3 Our measure of linguistic distance does not distinguish between Dutch and Flemish.

https://doi.org/10.1016/j.respol.2024.104971
Received 20 February 2023; Received in revised form 19 September 2023; Accepted 18 January 2024

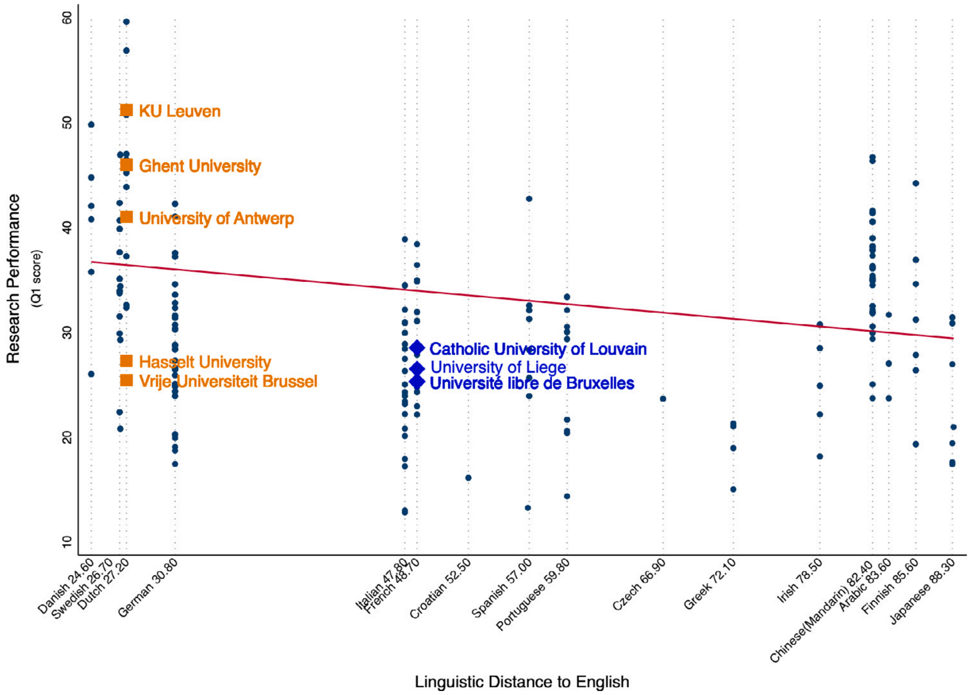

Fig. 1. Uncontrolled relationship between Linguistic Distance and Research Performance.
The figure scatter-plots the 2020 GRAS universities' linguistic distance of their regions' native language to English (horizontal axis) against their research performance (i.e., average Q1 score across social sciences subjects) in that same year (vertical axis), where Q1 score is the number of papers published by a university in journals with a first quartile journal impact factor for the academic subject (Shanghai Ranking, 2020). Universities in Belgium's Dutch-speaking region are highlighted in orange, while those in the French-speaking one are in blue. The red uncontrolled regression line captures a negative relationship between linguistic distance and social science research performance across all languages in our sample.

science Q1 scores than their counterparts in the French ($LD = 48.70$, average social science $Q1 = 26.9$) speaking region.[3] The red uncontrolled regression line illustrates this negative relationship between linguistic distance and social science research performance across all languages in our sample. Applied to the same two languages but controlling for university factors like full-time equivalent (FTE) staff, proportion of international staff, and number of students, as well as country-level factors like university degree rate, economic development, youth academic achievement, politics, culture, geographic distance to and trade with English speaking countries, among other factors, universities in Dutch speaking regions perform 12% better than those in French speaking ones across all academic subjects, not just social science.

Several mechanisms might be at work. First, researchers with a mother tongue further from the *lingua franca* may find acquiring the fluency required to publish in or read and build upon research at the frontier more difficult. They face greater language barriers to cooperate internationally and publish high-quality articles in English. They may find it more difficult to access resources through international co-affiliation networks (Hottenrott et al., 2021). Further, among articles from non-English speaking countries, those in the local language are less cited than English ones (Van Raan et al., 2011). Their published results may be less recognized globally, either because they appear in less prestigious outlets, they are less readable, or are of lower quality. There are several famous examples for the former, e.g., the original paper in Japanese on the Morishima elasticity of substitution in economics was only recognized after several American authors had independently discovered the concept (Morishima, 1967; Blackorby and Russell, 1981).

The remainder of the paper is organized as follows: The next section relates our contribution to the literature. Section 3 describes our empirical framework and data. Section 4 presents the results. Section 5 concludes with policy and managerial implications.

Appendices are available online: Appendix A provides summary statistics and correlation tables for samples other than those used in our preferred regression specifications, as well as our list of universities, their languages and linguistic distances to English. Appendix B presents additional robustness checks. Appendix C presents our heterogeneity analysis details in full, while Appendix D does the same for our placebo tests with other focal languages besides English.

2. Related literature and contribution

We contribute to three broad literatures: (1) drivers of university research performance, (2) barriers to knowledge transfer, and (3) inequality of opportunity in academia.

A large literature analyses the drivers of university research performance. MacLeod and Urquiola (2021) show that top US research universities' incentives, like higher salaries, tenure criteria and lower teaching loads increase research performance. Educational policies affect research outputs, for instance, UK government policies that focus on improving the efficiency of teaching outputs negatively affect universities' research outputs (Glass et al., 1998). National funding drives university research too. Many countries use performance-based funding to improve universities' international competitiveness (Sörlin, 2007). Changes in funding sources affect US research universities' performance (Foltz et al., 2012). However, the direct productivity effect of such incentives is unclear; multi-level communication and working environment matter more (Anon, 2010). The operational efficiency and internal resource allocation among departments and faculties affect university performance (Naderi, 2022; Zharova et al., 2022). More research funding, university autonomy on hiring and pay decisions, as well as student and staff mobility all increase university research performance (Aghion et al., 2008). While controlling for these factors, we add a national endowment as a source of competitive advantage to this list of drivers: linguistic proximity to the *lingua franca*.

Although our results directly explain performance, especially as measured by research outputs, we also speak indirectly to the literature on knowledge transfer and diffusion, as these mechanisms likely lie behind our findings. Amano et al. (2023) quantify the negative impact of language barriers on academic career development for non-native English speakers, who spend more effort in disseminating scientific research in multiple languages. Yang et al. (2022) show that linguistic distance between the language of parent companies' countries and that of their foreign subsidiaries reduces the effectiveness of both face-to-face and remote communication in international business. Jacobson et al. (2004) discuss structural barriers to knowledge translation activities in universities and research institutions, indicating that promotion, tenure, resources and funding influence knowledge transfer. Geographic distance, visa requirements and border restrictions inhibit international scientific communication and knowledge transfer (Appelt et al., 2015; Orazbayev, 2017). Geographic distance negatively relates to patent citations (Drivas and Economidou, 2015; MacGarvie, 2005), but the effect decreases when controlling for personal ties between scholars (Head et al., 2019). Geographic distance also reduces knowledge spillovers through collaboration (van der Wouden and Youn, 2023). Geuna and Muscio (2009) argue that cultural distance, management experience, and researcher involvement affect knowledge transfer between universities and firms. Miguelez and Temgoua (2020) document greater knowledge flow between countries with a common official language. International trade, indirectly, relies on knowledge transfer, and the trade literature recognizes the importance of language as a barrier (Melitz, 2008). Our analysis suggests that "genetic" distance between different languages further impedes international knowledge diffusion in academia, even after controlling for correlates like cultural and geographic distance.

Finally, we contribute to a literature on inequality of opportunity in academia. Most previous studies focus on classical attributes of discrimination like gender and race (Thelwall and Mas-Bleda, 2020; Kim and Patterson, 2020; Viglione, 2020; Kim et al., 2022). A literature in linguistics hypothesizes that non-native English-speaking scholars and students are also disadvantaged (Politzer-Ahles et al., 2016; Flowerdew, 2019; Hyland, 2016). Indeed, Elder and Davies (1998) show that linguistic distance between mother tongue and English (our explanatory variable) negatively correlates to students' academic performance. Linguistic distance generates language acquisition hurdles, negatively affecting school performance (Galloway and Gjefsen, 2020). We quantitatively confirm that this disadvantage carries over to academic research performance. Van Raan et al. (2011) argue that citation-based measurements of research performance may discriminate against non-native English speakers, because non-English articles are less cited. Our results suggest that, even among non-native English speakers, the degree of discrimination can be expected to increase with the distance between mother tongue and English, and the effect also manifests in top journal output.

3. Empirical framework & data

3.1. Model specification

To identify the impact of language distance to English on research performance, we consider pooled regression models.[4] Since countries' linguistic distance was exogenously set before any inputs to Shanghai Ranking's GRAS were contemplated, the primary threat to causal identification comes from university research performance drivers that correlate to linguistic distance. Hence, we use the following controlled OLS regression model:

$$lnY_{sict} = \alpha + \beta lnLD_i + \boldsymbol{X}'\lambda + \eta_s + \theta_c + \delta_t + \epsilon_{sict}, \quad (1)$$

where s indexes subjects, i universities, c countries, and t years. Our dependent variable Y_{sict}, is university i's research performance in subject s, in country c, in year t. Our explanatory variable of interest is linguistic distance (LD). We hypothesize linguistic distance negatively affects universities' research performance: $\beta < 0$. We log both Y and LD, as the log–log form produces the best fit, as measured by adjusted R-squared (Stock and Watson, 2012). It also facilitates interpretation of the results.[5]

Our preferred estimates come from a pooled model due to its asymptotically lower standard errors (Wooldridge, 2015). It uses both the between and within variation for identification. Controls and corresponding (non-causal) coefficients are given by $\boldsymbol{X}$ and λ. We control for subject fixed effects η_s, country fixed effects θ_c (in a single specification),[6] and year fixed effects δ_t. We cluster on language to allow for unknown serial and cross-sectional correlation in ϵ_{sict}. Because linguistic distance does not vary with time, we assume that the error has no university-specific, time invariant component. Assuming, conditional on a set of controls, that the errors do not correlate with (historically fixed) linguistic distance, its coefficient may be interpreted causally.

3.2. Data and variables

Our data comes from Shanghai Ranking's *Global Ranking of Academic Subjects* (GRAS) in 2020 and 2021.[7] It includes 54 subjects across five fields: natural sciences, engineering, life sciences, medical sciences, and social sciences.[8] After omitting universities lacking data for the full-time equivalent (FTE) staff index and those in English-speaking countries (189), 310 universities across 33 non-English-speaking countries and 24 languages remain. In some specifications, we reintroduce the 189 universities in English-speaking countries, for which we have FTE staff index, to check for robustness. Table A.5 in Appendix lists all universities in our sample, their country and local language. We discuss each variable below.

3.2.1. Dependent variable

We measure research performance with the GRAS Q1 score, the number of papers published by a university in journals with a first quartile journal impact factor in an academic subject (Shanghai Ranking, 2020). It captures a university's highest quality research output in 54 academic subjects across natural sciences, engineering, life sciences, medical sciences, and social sciences. GRAS obtains publication data from Web of Science and InCites, which include publications in many languages; however, few, if any, Q1 journals are published in languages

[4] See Table B.1 for results using the between estimator, Tables B.2 and B.3 for results using cross-sectional regression models.

[5] See Table B.4 for results using level Y and quadratic LD in Appendix.

[6] While controlling for unobserved, time-invariant country level factors appeals theoretically, the only countries in our sample with multiple language regions are Belgium, Italy, and Switzerland, which raises questions of external validity for results from models with country fixed effects.

[7] We drop 2022, because it includes the first year of the pandemic.

[8] The 54 academic subjects are, by field, (1) natural sciences — mathematics, physics, chemistry, earth sciences, geography, ecology, oceanography, and atmospheric science; (2) engineering — mechanical engineering, electrical & electronic engineering, automation & control, telecommunication engineering, instruments science & technology, biomedical engineering, computer science & engineering, civil engineering, chemical engineering, materials science & engineering, nanoscience & nanotechnology, energy science & engineering, environmental science & engineering, water resources, food science & technology, biotechnology, aerospace engineering, marine/ocean engineering, transportation science & technology, remote sensing, mining & mineral engineering, and metallurgical engineering; (3) life sciences — biological sciences, human biological sciences, agricultural sciences, and veterinary sciences; (4) medical sciences — clinical medicine, public health, dentistry & oral sciences, nursing, medical technology, and pharmacy & pharmaceutical sciences; (5) social sciences — economics, statistics, law, political sciences, sociology, education, communication, psychology, business administration, finance, management, public administration, hospitality & tourism management, and library & information science.

other than English.[9] Q1 score in 2020 counts total papers published 2014–2018 in the academic subject, while Q1 score in 2021 counts papers published 2015–2019. We prefer the GRAS from Shanghai Ranking to other international rankings metrics, as it objectively measures research performance, and does not include subjective factors, like reputation, or non-research factors, like graduates' starting salaries.

3.2.2. Independent variable

Linguistic distance captures the lexical relatedness between English and the official language of university i's country or region. Following recent innovations in the linguistics literature (see e.g., Galloway and Gjefsen (2020) and Yang et al. (2022)), we pull our measure of linguistic distance from the freely available tool on eLinguistics.net (eLinguistics, 2020). Their website provides a full description of eLinguistics' methodology. The basic idea is that words for certain objects or concepts, like 'eye,' 'death,' 'two,' and 'name,' exist in all languages with stable and consistent meanings but that their pronunciation and spelling has evolved and branched slowly over time and migration. eLinguistics measures the average sound correspondence of the consonants in 18 of these carefully selected "genetic marker" words across languages. Exact sound matches for a consonant pair receive a score of 100, unrelated or extra sounds a score of 0, and related sounds a score in between as defined in Brown et al. (2013). The average correspondence is computed across the 18 marker words and differenced from 100 to yield their linguistic distances, or genetic proximities, on a scale from 0 to 100. Importantly, since linguistic distance is cardinal, we can take its logarithm; however, since $ln[0]$ is undefined, marginal effects in log-specifications are undefined for universities in English-speaking countries.[10] The average linguistic distance from universities' local language to English in our sample is 51.5 with a standard deviation of 23.7. Highly related languages score below 30, related languages between 30 and 50, remotely related languages between 50 and 78, while those with higher scores have no recognizable relationship. For reference, German has a linguistic distance of 30.8. As a verification of their algorithmic method, eLinguistics generates a language evolutionary tree fully automatically that closely matches that agreed upon by linguistic historians (eLinguistics, 2020).

Several other measures of linguistic distance exist in the economics and innovation literature. Research on international trade uses indicators for common language or cost of translation to quantify barriers to trade (Melitz, 2008). Melitz and Toubal (2014) add their own pairwise language proximity based on language tree data in Ethnologue (2009). Chiswick and Miller (2005) develop another novel measure based on the ease with which Americans gain proficiency in various foreign languages. Although we expect other linguistic distance metrics would generate qualitatively similar results, we choose eLinguistics' metric for its transparency, availability, establishment in the literature, and because it is continuous across many languages and can be interpreted at both the individual and national level.

3.2.3. Control variables

Full-time equivalent (FTE) staff positively correlates with language distance (ρ = 0.354) and positively correlates with research performance (ρ = 0.051). Asian universities like those in China with greater language distance to English tend to have relatively larger FTE staffs. Universities with greater academic staff size might produce more high-quality research. We generate the FTE staff index from Shanghai Ranking's *Academic Ranking of World Universities*.[11] Omitting it might understate our effect of interest.

International staff proportion negatively correlates with language distance (ρ = −0.126) and positively correlates with research performance (ρ = 0.184). Universities with local languages closer to English, especially those of higher quality, probably attract more international staff. Though one might argue this is an effect of linguistic distance, we conservatively control for it. Omitting it would potentially increase the magnitude of the hypothesized negative effect of linguistic distance on performance.

Number of students negatively correlates with language distance (ρ = −0.082) and positively correlates with research performance (ρ = 0.016). Universities with greater size could be more likely to produce high-quality research. Omitting it could inflate our expected negative effect.

University degree rate is the percentage of population aged 25 and over that has attained or completed a Bachelor's degree or equivalent. It positively correlates with language distance (ρ = 0.082) and positively correlates with research performance (ρ = 0.106). Countries with a higher degree rate probably produce more high-quality research. Omitting it could underestimate our expected negative effect.

Number of native speakers positively correlates with language distance (ρ = 0.171) and negatively correlates with research performance (ρ = −0.035). We include this to control for English facility within the language group — when the number of native speakers is large, so is the volume of media, and the relative value of learning English may be lower. Its omission could overstate our effect.

Number of English speakers, a proxy for English proficiency at the country level, negatively correlates with language distance (ρ = −0.124) and negatively correlates with research performance (ρ = −0.041). Countries with higher English proficiency face fewer language barriers in scientific communication. So, omitting it in the absence of our other controls might understate our effect. It is, however, positively correlated to population size, and linguistic proximity to English likely drives its English proficiency. So, its effect in the battery of controls likely makes our estimate of the effect of linguistic distance on research performance more conservative.

Programme for International Student Assessment (PISA) score, a measure of a country's 15-year-old pupils' academic performance in mathematics, science, and reading, proxies for the quality of a country's statutory education. It positively correlates with language distance (ρ = 0.240), because Asian pupils perform well in PISA. Intuitively, it positively relates to research performance (ρ = 0.162). Hence, omitting it could understate our effect of interest.

We control for social culture. **Democracy index**, a proxy for political (and economic) distance to English-speaking countries by the Economics Intelligence Unit (2020), negatively correlates with language distance (ρ = −0.504) and positively with research performance (ρ = 0.040). Hence, its omission would likely overstate our effect, but its inclusion also properly accounts for the effect of democracy in countries like Japan and South Korea, which are linguistically distant but politically and economically close to Western democracies. Relatedly, we add two proxies of cultural distance (Hofstede, 1984). **Power distance**, a measure of how hierarchical a country is, positively correlates with language distance (ρ = 0.494) and positively with research performance (ρ = 0.002). Its omission could induce an understating bias. **Long-term orientation** positively correlates with language distance (ρ = 0.050) and positively correlates with research performance (ρ = 0.057), its omission would potentially underestimate our effect. These proxies from Hofstede (1984) were measured in the 1980s, which might not

[9] Because we were concerned that Web of Science and InCites might overlook highly cited, non-English journals, we manually checked Google Scholar Metrics for omissions from the subject Q1 lists and could not identify any.

[10] For more on the effect of LD in English-speaking countries, refer to the dicussion preceding the results from a quadratic model in Table B.4.

[11] Shanghai Ranking provides FTE as an index, indicating the university's number of FTE staff relative to the baseline university, the California Institute of Technology, with FTE staff index of 1. A university with FTE staff index of 3 has three times as many FTE staff as the California Institute of Technology.

Table 1
Variable description.

Variable	Description	Source
Q1 score	The number of papers published by a university's subject in journals with Q1 journal impact factor quartile during a five-year period	www.shanghairanking.com
Linguistic distance (LD)	Language distance between English and a university's local language	www.elinguistics.net/Compare_Languages.aspx
Full-time equivalent (FTE) staff	An index showing a university's FTE staff number relative to the baseline university	www.shanghairanking.com
International staff proportion (ISP)	University's international staff number as a percentage of total staff	www.qs.com/rankings
Number of students	University's number of students	www.timeshighereducation.com
University degree rate	A country's cohort share obtaining at least a bachelor's degree	uis.unesco.org
Number of native speakers	Global population of native speakers	jakubmarian.com/european-languages-by-number-of-native-speakers
Number of English speakers	National population of English-as-a-second-language speakers	ec.europa.eu/eurostat
PISA 2009 score	Average science, mathematics and reading score of Programme for International Student Assessment (PISA) in 2009	www.oecd.org/pisa/data
Democracy index	Economist Intelligence Unit's (EIU) index (0–10) to measure democracy	www.eiu.com/topic/democracy-index/
Power distance	A country's attitude towards power inequality (0–100)	www.hofstede-insights.com/product/compare-countries/
Long-term orientation	A country's attitude towards present and future challenges (0–100)	www.hofstede-insights.com/product/compare-countries/
Population	A country's population	data.worldbank.org
GDP	A country's gross domestic product, measured in $US	data.worldbank.org
20-year GDP growth	A country's GDP per capita growth over the last 20 years	data.worldbank.org
Trade	A country's total bilateral trade value (measured in $US) with the UK, the USA, Australia, New Zealand, and Canada	stats.oecd.org
Geographic distance (GD) to English	Weighted geographic distance between the university's country and the UK, the USA, Australia, New Zealand, and Canada	www.cepii.fr/CEPII/en, data.worldbank.org

be the most up-to-date cultural indicators, but cultural factors are relatively stable.

Clearly, a country's size and economic development could both drive research performance, and they correlate to linguistic distance. **Population** positively correlates with language distance (ρ = 0.144) and negatively correlates with research performance (ρ = −0.097). Omitting it could overstate our effect of interest. **GDP** positively correlates with language distance (ρ = 0.013) and negatively with research performance (ρ = −0.065). Omitting it could overstate our effect. There is, however, reason to believe that the effect of research investment is cumulative, and that current GDP per capita may not completely reflect that investment. Hence, we also control for **Long-term GDP growth rate**. We use the growth rate of a country's GDP per capita (current U.S. dollars) over a 20-year period, e.g., GDP growth rate in 2019 = (GDP per capita in 2019 − GDP per capita in 2000) / GDP per capita in 2000. It positively correlates with language distance (ρ = 0.361) and negatively with research performance (ρ = −0.001). The research capabilities of rapidly developing countries, like China, are still catching up. Hence, omitting it could overstate our effect of interest.

Linguistic distance to English surely relates to economic integration with English speaking countries. Although, as previously mentioned linguistic distance is itself a barrier to international trade, we conservatively control for it. We define **Trade** as a country's total bilateral trade value with the United Kingdom (UK), the United States (USA), Australia, New Zealand, and Canada. It negatively correlates with language distance (ρ = −0.080) and has a weak negative correlation with research performance (ρ = −0.011). Omitting it could understate our effect. Though highly correlated to trade volume, we also control for **Geographic distance** (GD) to English-speaking countries (i.e., the weighted geographic distance to the UK, the USA, Australia, New Zealand, and Canada).[12] It positively correlates with linguistic distance (ρ = 0.772) and positively with research performance (ρ = 0.062). Geographic proximity might facilitate research cooperation through lower travel costs. Omitting it could introduce positive bias in our expected negative effect.

Table 1 summarizes the above variable descriptions and provides the data source for each variable. Q1 score, university degree rate, democracy index, population, GDP, long-term GDP growth rate and trade vary over time. As Q1 score in 2020 (2021) covers the published papers during 2014–2018 (2015–2019), for each of these time-varying control variables, we use its average value over the same time period, i.e., 2014–2018 (2015–2019). Linguistic distance to English, power distance, long-term orientation, and geographic distance are time-invariant. We assume FTE staff index, international staff proportion, number of students, number of native speakers, number of English speakers, and PISA score remain constant within our two-year sample period.[13] We use the 2009 PISA score, because by 2021 this pupil cohort was 26 years old, approximately when academic researchers first begin publishing. Table 1 describes our variables and provides their original sources. Table 2 displays summary statistics, and Table 3 presents correlations for all variables in our 2020 cross-section of the data, which correspond to those variables used our preferred, pooled specification.[14]

4. Results

Table 4 presents our regression results, using the pooled sample for 2020 and 2021. The baseline model (with fixed effects for subject and year) appears in column (1). Column (2) adds university characteristics: an index for full-time equivalent (FTE) staff number, international

[12] We weight distance by population: $GD\_English_i = \sum_{j \in EC}(Pop_j \times GD_{ij}) / \sum_{j \in EC} Pop_j$, where EC ={ USA, Canada, UK, Australia, New Zealand }.

[13] PISA scores update only triennially.

[14] Our preferred specification excludes universities in China and in English-speaking countries. However, summary statistics and correlation coefficients for data including universities in English speaking countries, both with and without universities in China can be found in Appendix A.

Table 2
Summary statistics (excluding China and English-speaking Countries).

Variable	Mean	Example at mean	SD	Min	Max
Q1 score	34.2	Radboud University Nijmegen - Public health	13.4	6.8	100.0
LD to English	51.5	French	23.7	24.6	96.3
FTE staff	4.0	University of Potsdam (Germany)	1.0	0.6	9.5
ISP (%)	21.5	Heidelberg University (Germany)	17.8	0.0	79.4
Students	27 023.1	University of Kentucky (USA)	15 468.2	539.0	119259.0
University degree rate (%)	26.0	Italy	5.5	13.2	37.6
Native speakers (millions)	91.1	Switzerland	106.5	1.7	470.0
English speakers (millions)	16.9	Italy	13.8	1.6	45.0
PISA 2009 score	501.3	Norway	28.3	347.0	545.7
Democracy index	8.2	Austria	0.9	1.9	9.9
Power distance	46.8	Hungary	16.4	11.0	95.0
Long-term orientation	63.7	France	20.3	20.0	100.0
Population (millions)	45.3	Spain	43.3	4.2	208.58
GDP (billions)	1607.0	South Korea	1380.1	51.4	4908.1
20-year GDP growth (%)	84.8	Spain	49.0	4.9	234.2
Trade (billions)	139.8	Netherlands	114.7	1.2	369.7
GD to English (thousands km)	7.0	Poland	1.9	5.2	12.1

Table 3
Correlation coefficients (excluding China and English-speaking Countries).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1) Q1 score	1.000																
(2) Language distance to English	−0.048**	1.000															
(3) FTE staff	0.051 ***	0.354 ***	1.000														
(4) International staff proportion (%)	0.184 ***	−0.126 ***	−0.459 ***	1.000													
(5)Number of total students	0.016	−0.082 ***	0.541 ***	−0.394 ***	1.000												
(6)University degree rate (%)	0.106 ***	0.082 ***	−0.166 ***	0.309 ***	−0.301 ***	1.000											
(7)Number of native speakers (millions)	−0.035*	0.171 ***	0.292 ***	−0.278 ***	0.334 ***	−0.350 ***	1.000										
(8)Number of English speakers (millions)	−0.041**	−0.124 ***	0.128 ***	−0.310 ***	0.037*	−0.014	0.027	1.000									
(9) PISA 2009 score	0.162 ***	0.240 ***	−0.092 ***	0.191 ***	−0.375 ***	0.425 ***	−0.363 ***	0.009	1.000								
(10) Democracy index	0.040**	−0.504 ***	−0.253 ***	0.016	−0.182 ***	0.134 ***	−0.290 ***	−0.028	0.396 ***	1.000							
(11) Power distance	0.002	0.494 ***	0.390 ***	−0.257 ***	0.286 ***	−0.119 ***	0.346 ***	−0.006	−0.117 ***	−0.754 ***	1.000						
(12) Long-term orientation	0.057 ***	0.050 ***	0.007	−0.136 ***	−0.141 ***	0.389 ***	−0.138 ***	0.640 ***	0.493 ***	0.032*	0.138 ***	1.000					
(13) Population (millions)	−0.097 ***	0.144 ***	0.295 ***	−0.516 ***	0.159 ***	−0.166 ***	0.271 ***	0.607 ***	−0.175 ***	−0.152 ***	0.228 ***	0.449 ***	1.000				
(14) GDP (millions US$)	−0.065 ***	0.013	0.214 ***	−0.412 ***	0.084 ***	−0.056 ***	0.136 ***	0.769 ***	0.058 ***	0.014	0.048**	0.597 ***	0.868 ***	1.000			
(15) Long-term GDP growth rate (%)	−0.001	0.361 ***	0.136 ***	0.083 ***	−0.120 ***	0.206 ***	−0.009	−0.038**	−0.001	−0.178 ***	0.155 ***	−0.039**	−0.291 ***	−0.437 ***	1.000		
(16) Trade (billions US$)	−0.011	−0.080 ***	0.034*	−0.164 ***	−0.070 ***	0.111 ***	0.022	0.881 ***	0.209 ***	0.083 ***	−0.073 ***	0.707 ***	0.674 ***	0.891 ***	−0.269 ***	1.000	
(17) GD to English (thousands km)	0.062 ***	0.772 ***	0.244 ***	0.049 ***	−0.159 ***	0.193 ***	0.091 ***	0.012	0.287 ***	−0.518 ***	0.438 ***	0.315 ***	0.231 ***	0.122 ***	0.249 ***	0.091***	1.000

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

staff proportion, and number of students. Column (3) adds country-level, educational, demographic, political, and cultural characteristics: university degree rate, number of native speakers, number of English speakers, 2009 PISA score, democracy index, power distance, and long-term orientation. Column (4) adds country-level economic factors: population, GDP, and long-term GDP growth rate. Column (5) adds a country's total bilateral trade value with English-speaking countries (i.e., the UK, the USA, Australia, New Zealand, and Canada), and the weighted geographic distance to the above English-speaking countries. Column (6) is our preferred model. It excludes China to reduce bias from, for instance, over-representation of Chinese universities, from the fact that the Shanghai Ranking is Chinese, and that Chinese scholars disproportionately publish in Chinese journals. It shows that the effect is robust to excluding Chinese universities, which make up 17% of universities in non-English-speaking countries in our sample. Column (7) adds country fixed effects. Although these controls for unobserved, time-invariant country-level factors appeal theoretically, the number of countries with top universities in multiple local language regions is few. The identification in column (7) essentially comes only from Belgium, Italy, and Switzerland. In models (3) through (7) the coefficient on LD is statistically significant at the 0.1% level.

More importantly, the effect is economically significant, arguably even large. Our preferred model in column (6) indicates that a 1% increase in linguistic distance to English decreases a university's subject Q1 score by 0.215%.[15] Subject to the aforementioned caveat, controlling for unobserved, time-invariant country-level factors, the magnitude increases to 0.360% in column (7). More tangibly, Belgian universities in French ($LD = 48.70$) speaking regions achieve between 12% (using column (6)) and 19% (using column (7)) lower subject-level Q1 scores than those in Dutch ($LD = 27.20$) ones, holding all else constant.

Although causal interpretation of the control variables' coefficients is unwarranted, it suggests that higher faculty to student ratio (i.e., FTE staff, holding Students constant), larger international staff proportions, and higher PISA 2009 scores (i.e., higher faculty ability) are associated with better university research performance.

The above regression results show that language distance to English negatively affects universities' research performance. Apparently, the more different their mother tongue is from English, the harder it is for scholars to produce and publish high-quality research.

Exploring mechanisms, we check whether the effect of language distance differs across fields. We rerun the models of Table 4 for five different academic fields, restricting the sample to just observations in the focal field for each. The complete results, including coefficient estimates for the control variables, can be found in Appendix C. Table 5 summarizes these results — each row provides the effect of linguistic distance on Q1 score for our preferred model, standard error of the coefficient, and the number of subsample observations for one academic field. Language distance has the greatest impact on medical science output, falling gradually from life sciences to social sciences to engineering until the effect is statistically insignificant for natural sciences. We know of no rank ordering of these fields by the relative importance of verbal communication in them, but it is intuitive that poor English will be less penalized in fields where ideas are more readily communicated in the universal language of mathematics.

Finally, placebo tests show that only linguistic distance to English matters. Repeating our main analysis for Korean, Japanese, Estonian,

[15] In a log–log model, the relationship between a ΔX change in X and a ΔY change in Y is given by $\beta = \ln[1 + \Delta Y/Y]/\ln[1 + \Delta X/X]$. This implies that a (small) 1% change in X is associated with approximately a β% change in Y.

Table 4
Influence of language distance on research performance.

	Dep. Var. = Q1 score (*ln*)						
Variables	Pooled 2020–2021						
						excl. China	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LD to English (*ln*)	0.036	−0.049	−0.185***	−0.193***	−0.252***	−0.215**	−0.360***
	(0.073)	(0.047)	(0.042)	(0.043)	(0.051)	(0.058)	(0.062)
FTE staff (*ln*)		0.368***	0.235***	0.194***	0.140*	0.134*	0.041
		(0.066)	(0.045)	(0.051)	(0.053)	(0.056)	(0.055)
ISP (%)		0.004***	0.003***	0.005**	0.004**	0.006***	0.005***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Students (*ln*)		0.010	0.038	0.051	0.072*	0.081**	0.118***
		(0.034)	(0.030)	(0.032)	(0.026)	(0.027)	(0.031)
University degree rate (%)			0.002	0.005	0.004	0.005	1.922***
			(0.003)	(0.003)	(0.002)	(0.002)	(0.259)
Native Speakers (*ln* millions)			0.018	0.013	0.019	0.021*	3.389***
			(0.012)	(0.013)	(0.010)	(0.010)	(0.610)
English Speakers (*ln* millions)			−0.008	0.001	0.023	0.061	20.067***
			(0.019)	(0.034)	(0.025)	(0.046)	(2.109)
PISA 2009 score (hundreds)			0.307*	0.360**	0.312**	0.289*	67.678***
			(0.121)	(0.104)	(0.098)	(0.114)	(7.164)
Democracy index			−0.028	0.007	0.032	0.057	−5.748***
			(0.018)	(0.024)	(0.021)	(0.030)	(0.692)
Power distance			0.001	0.002	0.003*	0.003*	−0.910***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.106)
Long-term orientation			−0.001	−0.002	−0.003**	−0.003*	−1.804***
			(0.002)	(0.001)	(0.001)	(0.001)	(0.195)
Population (*ln* millions)				0.130	0.067	0.098	1.423
				(0.072)	(0.077)	(0.066)	(1.040)
GDP (*ln* millions US$)				−0.109	−0.048	−0.095	−0.472
				(0.067)	(0.090)	(0.098)	(0.425)
Long-term GDP growth rate				0.007	0.024	0.003	−0.062
				(0.013)	(0.013)	(0.045)	(0.122)
Trade (*ln* billions US$)					−0.012	−0.021	0.016**
					(0.038)	(0.041)	(0.005)
GD to English (thousands km)					0.035***	0.037**	4.363***
					(0.009)	(0.010)	(0.495)
Constant	3.590***	3.168***	2.156***	2.571***	1.826*	1.752	−272.241***
	(0.258)	(0.364)	(0.550)	(0.522)	(0.778)	(0.860)	(29.578)
N	11 562	11 562	11 562	11 562	11 562	9246	9246
Subject FEs	✓	✓	✓	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓	✓	✓	✓
Country FEs							✓
R-squared	0.293	0.361	0.396	0.400	0.405	0.406	0.427
Adjusted R-squared	0.290	0.358	0.392	0.396	0.401	0.401	0.421

[1] * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, standard errors in parentheses and clustered at the university's local language.
[2] The dependent variable is the logged Q1 score (0–100) capturing a university's research performance by subject. The variable of interest is the logged linguistic distance between English and the university's local language (0–100).
[3] (1) is the baseline model. (2) adds the university's relative full-time equivalent staff number, international staff proportion, and the university's number of students. (3) adds university degree rate in a country, the number of worldwide native speakers of the university's local language, the number of English-as-a-second-language speakers, the country's average PISA 2009 score, democracy index, power distance, and long-term orientation index. (4) adds the country's population, GDP, and GDP per capita growth rate in a 20-year period. (5) adds trading value with the US, UK, Australia, New Zealand, and Canada, and the weighted geographic distance to the US, UK, Australia, New Zealand, and Canada. (6) is the same model as (5) excluding China. (7) is the same model as (6), controlling for country fixed effects (FEs). Pooled regressions control for subject FEs and year FEs.

and Arabic (all very distant to English) shows positive effects of distance to these languages on research performance, while repeating it for French (moderately distant from English) shows no effect (see Table 6). The controls for geographic distance and trade in these placebo tests are taken relative to and with South Korea, Japan, Estonia, Saudi Arabia, and France, respectively. Coefficients in Table 6 are taken from columns (6) of Tables D.1 to D.5. This bolsters our interpretation that the effects of distance from English speaking countries on research performance that we observe are, indeed, operating through the channel of language.

5. Conclusion

We show that universities in regions with languages more different from English exhibit worse research performance. Furthermore, the

Table 5
Influence of linguistic distance across fields (excl. China).

Field	LD Coefficient	SE	Observations
Medical sciences	−0.273***	0.066	1371
Life sciences	−0.250*	0.099	963
Social sciences	−0.226*	0.090	1827
Engineering	−0.198***	0.050	3253
Natural sciences	−0.125	0.068	1832

[1] * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, standard errors (SE) clustered at the university's local language.
[2] For each row, the model is the same as column (6) in Table 4, but the data is restricted to observations in the corresponding academic field. The panel is 2020–2021. The dependent variable is subject-level *ln[Q1]* from the Shanghai Ranking's GRAS, while the coefficient corresponds to the impact of linguistic distance to English by academic field.
[3] The sample size differs among fields, as different universities and numbers of them are ranked for each field.

Table 6
Influence of distance to other languages.

Language	LD to English	LD coefficient	SE
Korean	90.0	0.467	0.585
Japanese	88.3	1.716 ***	0.242
Estonian	84.9	0.107 **	0.037
Arabic	83.6	0.529	0.844
French	48.7	−0.0002	0.049

[1] * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, standard errors (SE) clustered at the university's local language.
[2] The *LD coefficient* for each language, taken from columns (6) of Tables D.1 to D.5, indicates linguistic distance to other languages' effect on research performance.

effect is large: a 1% increase in language distance to English, decreases a university's Q1 score by 0.215%, in controlled regressions. When we add country fixed effects to hold constant time-invariant country-level factors we identify a 0.360% decrease across a limited number of countries with multiple language regions. More plainly, Belgian universities in French ($LD = 48.70$) speaking regions achieve between 12% and 19% lower subject-level Q1 scores than those in Dutch ($LD = 27.20$) ones, holding all else constant. Hence, the research playing field for nations, universities, and individual researchers is uneven.

What can be done to level it? Unfortunately, there is no reasonable policy lever which can change a country's or individual's linguistic distance to English. However, other research suggests that enhancing early primary school education in English improves facility with the language and academic performance later in life (Taylor and von Fintel, 2016). Many global universities are now switching language of instruction to English, and are seeing research performance improve as a result (Cao et al., 2022). This surely internationalizes faculties, improves their English skills, and reduces academic inbreeding (Seeber and Mampaey, 2022).

Of course, such policy changes have costs. More English training would crowd out other subjects. Several governments, e.g., in Belgium and China, now restrict the amount of teaching in English to address local employers' demands for local language skills and domestic educational inequalities. Besides, we detect such sizable effects despite the fact that most researchers today already have English as at least their second language, despite high researcher mobility and increasing numbers of universities that teach in English. So, while investments in childhood English education and internationalization of university teaching and staff may temper the effects of linguistic distance on research performance, understanding how effective these measures are, is the subject of future research. In the meantime, we hope that our first quantification of the research performance penalties induced by linguistic distance from the *lingua franca* will inform policy makers who must balance these trade-offs. In addition to manipulating English proficiency, it is reasonable to ask whether the metrics we use to evaluate universities' and individual researchers' performance should be calibrated, or include a broader range of factors, less directly influenced by English facility.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.respol.2024.104971.

References

Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., Sapir, A., 2008. Higher Aspirations: An Agenda for Reforming European Universities. Bruegel Blueprint Series, Volume 5, July 2008. Bruegel, https://www.bruegel.org/book/higher-aspirations-agenda-reforming-european-universities.

Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., González-Trujillo, J.D., Montaño-Centellas, F., Paudel, K., White, R.L., et al., 2023. The manifold costs of being a non-native English speaker in science. PLoS Biol. 21 (7), e3002184, https://doi.org/10.1371/journal.pbio.3002184.

Ammon, U., 2012. Linguistic inequality and its effects on participation in scientific discourse and on global knowledge accumulation–with a closer look at the problems of the second-rank language communities. Appl. Linguist. Rev. 3 (2), 333–355, https://doi.org/10.1515/applirev-2012-0016.

Anon, 2010. University research funding and publication performance—An international comparison. Res. Policy 39 (6), 822–834, https://doi.org/10.1016/j.respol.2010.03.003.

Appelt, S., van Beuzekom, B., Galindo-Rueda, F., de Pinho, R., 2015. Which factors influence the international mobility of research scientists? In: Global Mobility of Research Scientists. Elsevier, pp. 177–213, https://doi.org/10.1016/B978-0-12-801396-0.00007-7.

Baruffaldi, S.H., Landoni, P., 2016. Mobility intentions of foreign researchers: The role of non-economic motivations. Ind. Innov. 23 (1), 87–111, https://doi.org/10.1080/13662716.2015.1126502.

Bercovitz, J.E., Feldman, M.P., 2007. Fishing upstream: Firm innovation strategy and university research alliances. Res. Policy 36 (7), 930–948, https://doi.org/10.1016/j.respol.2007.03.002.

Blackorby, C., Russell, R.R., 1981. The morishima elasticity of substitution; symmetry, constancy, separability, and its relationship to the Hicks and Allen elasticities. Rev. Econom. Stud. 48 (1), 147–158, https://www.jstor.org/stable/2297127.

Brown, C.H., Holman, E.W., Wichmann, S., 2013. Sound correspondences in the world's languages. Language 4–29, https://www.jstor.org/stable/23357720.

Cao, Y., Triebs, T.P., Tumlinson, J., 2022. Language as a strategic choice — Drawing global research talent by switching to English. https://ssrn.com/abstract=4215942.

Chiswick, B.R., Miller, P.W., 2005. Linguistic distance: A quantitative measure of the distance between English and other languages. J. Multiling. Multicult. Dev. 26 (1), 1–11, https://doi.org/10.1080/14790710508668395.

Cohen, W.M., Nelson, R.R., Walsh, J.P., 2002. Links and impacts: The influence of public research on industrial R&D. Manage. Sci. 48 (1), 1–23, https://www.jstor.org/stable/822681.

Conti, A., Gaule, P., 2011. Is the US outperforming Europe in university technology licensing? A new perspective on the European Paradox. Res. Policy 40 (1), 123–135, https://doi.org/10.1016/j.respol.2010.10.007.

Di Bitetti, M.S., Ferreras, J.A., 2017. Publish (in English) or perish: The effect on citation rate of using languages other than English in scientific publications. Ambio 46 (1), 121–127, https://link.springer.com/article/10.1007/s13280-016-0820-7.

Drivas, K., Economidou, C., 2015. Is geographic nearness important for trading ideas? Evidence from the US. J. Technol. Transfer 40 (4), 629–662, https://link.springer.com/article/10.1007/s10961-014-9360-0.

Economics Intelligence Unit, 2020. Democracy index 2019: A year of democratic setbacks and popular protest. https://www.eiu.com/topic/democracy-index/.

Elder, C., Davies, A., 1998. Performance on ESL examinations: Is there a language distance effect? Lang. Educ. 12 (1), 1–17, https://doi.org/10.1080/09500789808666736.

eLinguistics, 2020. Quantifying the genetic proximity between languages. http://www.elinguistics.net/Compare_Languages.aspx.

Ethnologue, 2009. Ethnologue: Languages of the World, sixteenth ed. Summer Institute of Linguistics, International Academic Bookstore, Dallas, TX.

Flowerdew, J., 2019. The linguistic disadvantage of scholars who write in English as an additional language: Myth or reality. Lang. Teach. 52 (2), 249–260, https://doi.org/10.1017/S0261444819000041.

Foltz, J.D., Barham, B.L., Chavas, J.-P., Kim, K., 2012. Efficiency and technological change at US research universities. J. Product. Anal. 37, 171–186, https://www.jstor.org/stable/43549651.

Galloway, T.A., Gjefsen, H.M., 2020. Assimilation of immigrants: Does earlier school exposure matter? Econ. Educ. Rev. 76, 101976, https://doi.org/10.1016/j.econedurev.2020.101976.

Geuna, A., Muscio, A., 2009. The governance of university knowledge transfer: A critical review of the literature. Minerva 47 (1), 93–114, https://link.springer.com/article/10.1007/s11024-009-9118-2.

Geuna, A., Nesta, L.J., 2006. University patenting and its effects on academic research: The emerging European evidence. Res. Policy 35 (6), 790–807, https://doi.org/10.1016/j.respol.2006.04.005.

Giannopoulou, E., Barlatier, P.-J., Pénin, J., 2019. Same but different? Research and technology organizations, universities and the innovation activities of firms. Res. Policy 48 (1), 223–233, https://doi.org/10.1016/j.respol.2018.08.008.

Glass, J.C., McKillop, D.G., O'Rourke, G., 1998. A cost indirect evaluation of productivity change in UK universities. J. Product. Anal. 10, 153–175, https://www.jstor.org/stable/41769956.

Hamel, R.E., 2007. The dominance of English in the international scientific periodical literature and the future of language use in science. Aila Rev. 20 (1), 53–71, https://eric.ed.gov/?id=EJ1069221.

Head, K., Li, Y.A., Minondo, A., 2019. Geography, ties, and knowledge flows: Evidence from citations in mathematics. Rev. Econ. Stat. 101 (4), 713–727, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3163679.

Hermanu, A.I., Sari, D., Sondari, M.C., Dimyati, M., 2022. Is it necessary to evaluate university research performance instrument? Evidence from Indonesia. Cogent Soc. Sci. 8 (1), 2069210, https://doi.org/10.1080/23311886.2022.2069210.

Hofstede, G., 1984. Culture's Consequences: International Differences in Work-Related Values, vol. 5, Sage.

Hottenrott, H., Rose, M.E., Lawson, C., 2021. The rise of multiple institutional affiliations in academia. J. Assoc. Inf. Sci. Technol. 72 (8), 1039–1058, https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.24472.

Hyland, K., 2016. Academic publishing and the myth of linguistic injustice. J. Second Lang. Writ. 31, 58–69, https://doi.org/10.1016/j.jslw.2016.01.005.

Jacobson, N., Butterill, D., Goering, P., 2004. Organizational factors that influence university-based researchers' engagement in knowledge transfer activities. Sci. Commun. 25 (3), 246–259, https://doi.org/10.1177/1075547003262038.

Jaffe, A.B., 1989. Real effects of academic research. Am. Econ. Rev. 957–970, https://www.jstor.org/stable/1831431.

Kim, J., Lee, S.J., Marschke, G., The influence of university research on industrial innovation. Working Paper Series (11447). National Bureau of Economic Research http://dx.doi.org/10.3386/w11447.

Kim, E., Patterson, S., 2020. The pandemic and gender inequality in academia. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3666587.

Kim, L., Smith, D.S., Hofstra, B., McFarland, D.A., 2022. Gendered knowledge in fields and academic careers. Res. Policy 51 (1), 104411, https://doi.org/10.1016/j.respol.2021.104411.

Laursen, K., Salter, A., 2004. Searching high and low: What types of firms use universities as a source of innovation? Res. Policy 33 (8), 1201–1215, https://doi.org/10.1016/j.respol.2004.07.004.

MacGarvie, M., 2005. The determinants of international knowledge diffusion as measured by patent citations. Econom. Lett. 87 (1), 121–126, https://doi.org/10.1016/j.econlet.2004.09.011.

MacLeod, W.B., Urquiola, M., 2021. Why does the United States have the best research universities? Incentives, resources, and virtuous circles. J. Econ. Perspect. 35 (1), 185–206, https://www.aeaweb.org/articles?id=10.1257/jep.35.1.185.

Melitz, J., 2008. Language and foreign trade. Eur. Econ. Rev. 52 (4), 667–699, https://doi.org/10.1016/j.euroecorev.2007.05.002.

Melitz, J., Toubal, F., 2014. Native language, spoken language, translation and trade. J. Int. Econ. 93 (2), 351–363, https://doi.org/10.1016/j.jinteco.2014.04.004.

Metcalfe, J.S., 2010. University and business relations: Connecting the knowledge economy. Minerva 48 (1), 5–33, https://ideas.repec.org/p/cbr/cbrwps/wp395.html.

Miguelez, E., Temgoua, C.N., 2020. Inventor migration and knowledge flows: A two-way communication channel? Res. Policy 49 (9), 103914, https://doi.org/10.1016/j.respol.2019.103914.

Morishima, M., 1967. A few suggestions on the theory of elasticity. Keizai Hyoron (Economic Review) 16 (1), 144–150.

Naderi, A., 2022. Efficiency measurement of higher education units using multilevel frontier analysis. J. Product. Anal. 57 (1), 79–92, https://link.springer.com/article/10.1007/s11123-021-00621-0.

Orazbayev, S., 2017. International knowledge flows and the administrative barriers to mobility. Res. Policy 46 (9), 1655–1665, https://doi.org/10.1016/j.respol.2017.08.001.

Politzer-Ahles, S., Holliday, J.J., Girolamo, T., Spychalska, M., Berkson, K.H., 2016. Is linguistic injustice a myth? A response to Hyland. J. Second Lang. Writ. 34, 3, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5502761/.

Rosenberg, N., Nelson, R.R., 1994. American universities and technical advance in industry. Res. Policy 23 (3), 323–348, https://doi.org/10.1016/0048-7333(94)90042-6.

Seeber, M., Mampaey, J., 2022. How do university systems' features affect academic inbreeding? Career rules and language requirements in France, Germany, Italy and Spain. Higher Educ. Q. 76 (1), 20–35, https://doi.org/10.1111/hequ.12302.

Shanghai Ranking, 2020. Global ranking of academic subjects (GRAS) 2019. http://www.shanghairanking.com.

Sörlin, S., 2007. Funding diversity: Performance-based funding regimes as drivers of differentiation in higher education systems. Higher Educ. Policy 20 (4), 413–440, https://link.springer.com/article/10.1057/palgrave.hep.8300165.

Stock, J.H., Watson, M.W., 2012. Introduction to Econometrics, vol. 3, Pearson New York.

Taylor, S., von Fintel, M., 2016. Estimating the impact of language of instruction in South African primary schools: A fixed effects approach. Econ. Educ. Rev. 50, 75–89, https://doi.org/10.1016/j.econedurev.2016.01.003.

Tether, B.S., Tajar, A., 2008. Beyond industry–university links: Sourcing knowledge for innovation from consultants, private research organisations and the public science-base. Res. Policy 37 (6–7), 1079–1095, https://doi.org/10.1016/j.respol.2008.04.003.

Thelwall, M., Mas-Bleda, A., 2020. A gender equality paradox in academic publishing: Countries with a higher proportion of female first-authored journal articles have larger first-author gender disparities between fields. Quant. Sci. Stud. 1 (3), 1260–1282, https://doi.org/10.1162/qss_a_00050.

van der Wouden, F., Youn, H., 2023. The impact of geographical distance on learning through collaboration. Res. Policy 52 (2), 104698, https://doi.org/10.1016/j.respol.2022.104698.

Van Raan, A.F., Van Leeuwen, T.N., Visser, M.S., 2011. Severe language effect in university rankings: Particularly Germany and France are wronged in citation-based rankings. Scientometrics 88 (2), 495–498, https://link.springer.com/article/10.1007/s11192-011-0382-1.

Viglione, G., 2020. Are women publishing less during the pandemic? Here's what the data say. https://www.nature.com/articles/d41586-020-01294-9.

Wooldridge, J.M., 2015. Introductory Econometrics: A Modern Approach. Cengage Learning.

Yang, J.Y., Wen, L., Volk, S., Lu, J.W., 2022. Temporal boundaries and expatriate staffing: Effects of parent–subsidiary work-time overlap. J. World Bus. 57 (6), 101367, https://doi.org/10.1016/j.jwb.2022.101367.

Zharova, A., Härdle, W.K., Lessmann, S., 2022. Data-driven support for policy and decision-making in university research management: A case study from Germany. European J. Oper. Res. https://doi.org/10.1016/j.ejor.2022.10.016.